National Tsing Hua University Fall 2023 11210IPT 553000 Deep Learning in Biomedical Optical Imaging Report

AUTHOR ONE¹ 張皓旻

Student ID:112022533

Report:癌症組織分類以及 Data augmentation 的測試

在這份報告,我將進行癌症組織分類以及 Data augmentation 的實際運用,並且 測試不同程度的 Data augmentation 對結果的影響。

模型建立與訓練參數設定:

這份報告我將使用訓練好的模型 resnet18 來進行,搭配的權重為 IMAGENET1K_V1,並設定為可以調整模型所有參數,之後的全連接層為128個神 經元最後分成6類不同的癌症,如下圖Fig.1所示。

Fig.1 本次報告所使用的模型

並且其餘 hyperparameter 都設定相同, (CrossEntropyLoss、batch_size=32、epochs=30、optimizer=Adam(lr=1e-3)、StepLR(step_size=10, gamma=0.1))。

Data augmentation 的設定:

本次實驗將會測試三種程度的 Data augmentation,第一種是原始的訓練資料,沒有經過任何操作,第二種是輕度的 Data augmentation,只會將圖片進行水平翻轉、垂直翻轉以及旋轉這三種操作,如下圖 Fig.2 所示,第三種是大幅的 Data

augmentation,在第二種的操作上再加上其他種隨機的方法,如下圖 Fig.3 所示,可看到影像變化程度高於第二種。

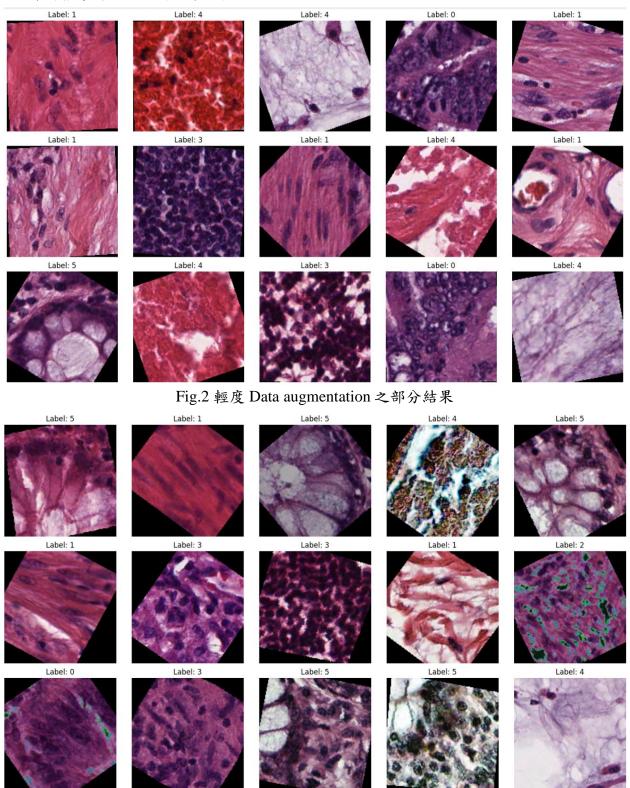


Fig.3 大幅的 Data augmentation 之部分結果

結果:

1. 原始的訓練資料:

結果如下圖 Fig.4 所示,可以發現其 overfitting 得非常嚴重,train acc 到達 100%,但 val acc 只有接近 95%,並且對於 test data 的正確率有 94.6%,雖然算是蠻高的正確率了,但也許 Data augmentation 之後可以改善 overfitting,並且得到更 general 的模型。

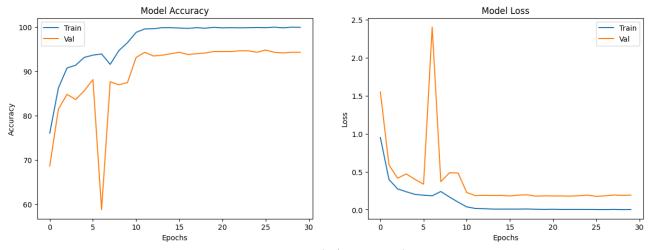


Fig.4 原始訓練資料的訓練結果

2. 輕度 Data augmentation:

結果如下圖 Fig.5 所示, overfitting 的現象改善了非常多,train acc 到達 96.7%,val acc 為 96.17%,十分接近,並且對於 test data 的正確率有 94.8%,雖然正確率沒有高非常多,但這模型應該是更 general 的,如果有更多種的 test data 就能驗證此猜想。

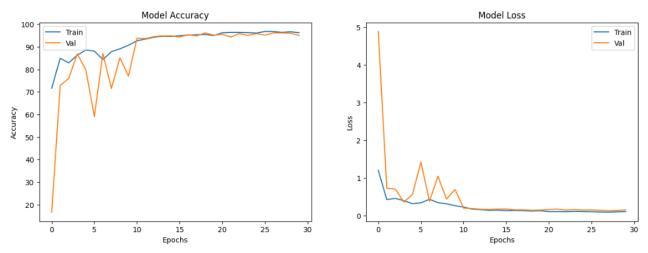


Fig.5 輕度 Data augmentation 的訓練結果

3. 大幅的 Data augmentation:

結果如下圖 Fig.6 所示, 同樣的 overfitting 改善非常多,train acc 到達約92.7%,val acc 為93%,對於 test data 的正確率有93.8%,正確率有些微下降,看來大幅的 Data augmentation 難度有點太高,可能要再訓練更多輪,或調整其他參數,不過也很好的改善了 overfitting,並且這會是比輕度 Data augmentation 情況下更 general 的模型。

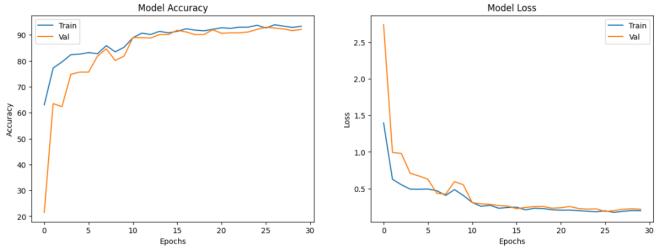


Fig.5 大幅 Data augmentation 的訓練結果